DOCUMENT RESUME

ED 126 566 EA 008 508

AUTHOR Doherty, Victor
TITLE Developments in Goal Based Measurement in the Portland Public Schools.
PUB DATE 9 Apr 76
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; *Educational Objectives; Elementary Secondary Education; Item Analysis; *Item Banks; Measurement Techniques; *Test Construction; *Testing
IDENTIFIERS *Portland Oregon Public Schools

ABSTRACT

For almost 20 years, the Portland public schools have maintained a testing program anchored in locally constructed tests. The district's newly developed tests are called goal-referenced tests, a version of criterion-referenced tests, that meet the district's purposes because they reflect what well-qualified teachers in the district believe should be taught and measured. In constructing the tests, curriculum personnel select goals they believe the test should measure; teachers develop items that measure those goals; the tests are given a trial administration and items are analyzed; the items, formats, and directions are revised; and the tests are administered for their intended use. The Rasch test scaling procedure that involves the identification of an equal-interval scale of difficulty for a given set of items based on information about item difficulty and total test performance for the group tested is used. The procedure yields information on item difficulty and an estimate of the ability of individuals and groups tested. This permits establishment of a scale that is independent of the norming population and allows for the creation of item pools. (Author/IRT)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED126566

4-9-76

Developments in Goal Based Measurement in the Portland Public Schools

Dr. Victor Doherty
Portland Public Schools

For almost 20 years the Portland Public Schools has maintained a testing program anchored in locally constructed tests and local norms using standard scores with a mean of 50 and a standard deviation of 10. Under the leadership of George Ingebo and Dean Forbes, this program served the purposes of research and evaluation in the school system much better, in our judgment, than publisher's tests with their elusive national norms and non-equal interval derived scores. It is a tribute both to the leadership of measurement people in the District and to the Superintendents under whom they served that such a program survived the pressures that constantly urged return to politically attractive uses of standardized tests and grade-equivalent scores.

Events of the past five years, however, have imposed on the Portland schools a need to advance its program to yet another stage of development; one which we hope will represent important progress in public school measurement.

An event in 1970-71 that helped precipitate change was the creation of three sub-districts having considerable autonomy in planning and evaluation. A Central Evaluation Department was created to monitor progress in each of the three new sub-districts, and to audit the evaluations that

EA 008 508

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

each was staffed to perform. Two areas chose to continue testing programs similar to the former district-wide program; the third chose to use nationally standardized tests with grade-level equivalents. In this third area, different test forms were also administered to students of different ability.

Problems in central auditing of sub-district evaluations were posed immediately by the different testing programs of the three areas. To preserve some elements of a common measurement base, city-wide administration of math and reading tests formerly used was required for grades 4 and 8, along with TAP math and reading tests for grade 11.

This much knowledge of the background of Portland's testing program is important in understanding what follows.

Shortly after the Central Evaluation Department was created in 1970-71, it became evident that program evaluation of the type desired in Portland simply could not occur without well-defined learning outcomes in the various courses of study. Behavioral objectives, with their extreme specificity and stated conditions of performance did not seem to be a viable type of outcome statement for use in planning and evaluating instructional programs. So we set about to create a type of statement that served these purposes effectively, and came up with something called a "course goal," which is simply a concise, clear statement of desired learning.

The Central Evaluation Department organized a three-county effort to develop this new tool for planning and evaluation. Over a four-year period, comprehensive, carefully classified sets of course goals were produced in 12 fields of study.

The tri-county goal-defining effort was intended to place a resource in the hands of teachers and administrators that would permit them to select rather than create statements of desired learning. This seemed necessary since attempts of school systems throughout the country to have teachers create such statements seemed to produce results of insufficient quality for successful planning and evaluation. The 12 course-goal collections created by the tri-county cooperative effort now provide a base for planning and measurement that is comprehensive and of acceptable quality.

Through the work of Fred Forster, we now have the ability to print out item results for each goal represented in each test developed for use in the system. In developing new tests, the first step is for curriculum personnel to select the goals they believe the test should measure. The second step is for teachers to develop items that measure those goals. In doing this, teachers follow a procedure for goal domain development published by the Northwest Evaluation Association. The third step is trial administration of test modules and item analysis. The fourth is revision of items, test formats, and directions based on item analysis

and experience with the trial administration; and fifth is administration of the test for its intended use in the system. At this point, information on extent of goal attainment is printed for each student and for each class for teacher use.

The resultant tests are what we term goal-referenced tests. They are simply another version of what are sometimes called objective-referenced or criterion-referenced tests. Their superiority for our purposes derives from the fact that they reflect what well qualified teachers in the District believe should be taught and measured.

I have not yet touched on a second development that reinforces this goal-referencing capability to open new testing potentials in Portland. The Rasch test scaling procedure, promoted by Ben Wright and others in this country, involves the identification of an equal-interval scale of difficulty for a given set of items based upon information about item difficulty and total test performance for the group tested. The Rasch procedure attempts to define item difficulty with the greatest precision possible on the basis of trial item administrations. The procedure can yield information on item difficulties for any test administered to any group; and also yields an estimate of the ability of individuals and groups tested.

What advantages does this method have over conventional test norming and scaling procedures? First, it permits establishment of a scale that is independent of a norming population. Given conditions of curricular

validity and good test construction, it appears that item calibrations based on administering a test to 200 or more students are very stable and quite robust with regard to the achievement level of the norm group.

A second advantage of the Rasch, and one of great importance to us, is the ability to create item pools through the administration of a large number of different tests, linked to one another by overlapping items. By obtaining difficulty values (calibrations) of the linking or overlapping items, and then adjusting the calibrations from one test to the other, it is possible to place all items in all tests on a difficulty continuum. The scale thus created makes it possible to secure comparable performance estimates for various groups attempting any items in the pool.

To understand the importance of this procedure it is necessary to return to our goal-based system of test construction. One of the persistent objections raised by teachers to measurement and especially to use of standardized tests, is the difficulty of finding or constructing tests that correspond to the outcomes sought by particular teachers. That objection can be overcome by a system that (1) permits teachers to select the goals they wish to have measured, (2) has calibrated items relating to those goals from which total-score estimates can be derived that are statistically comparable to those derived from any other set of items administered from the same pool.

The combined goal-referencing and Rasch scaling capabilities, if all assumptions, techniques, and procedures prove valid, should satisfy these two conditions.

Having a common metric for a large pool of items not only makes it possible to secure comparable measures for different groups working on different goals; it also makes possible the administration of simple tests to less able students and more difficult tests to more able students while retaining score-comparing and score-averaging capabilities.

Portland's test development work of the past two years has made increasing use of the capabilities just described. Following is a brief review of tests developed or under development in the school system:

Test	
City-wide reading and math tests, grades 4,8	Items in existing tests have been referenced to tri-county goals; goal attainment reports now provided teachers. Tests Rasch analyzed, but Rasch scaled scores not yet used.
Credit by examination tests for high school	Tests in 9th grade math, science and language arts have been created by teachers who wrote items to measure goals selected by teacher committees. Tests Rasch analyzed; but Rasch scale scores not yet used.
High school math course examinations	Over 150 forty-five minute modules with overlapping items have been created for measuring mid-term and second term achievement in 19 high school math courses. Mid-term tests Rasch analyzed; second term tests to be Rasch analyzed this summer. Program should be in standard use by mid-term, 1977, with Rasch scaled score reporting, possibly supplemented by standard scores (mean 50, S.D. 10). Results reported by goal as well as by total score.
Elementary "level tests" in mathematics and reading, grades 3-8	Over 2000 items Rasch calibrated in elementary reading and math through administration of modules in Portland and cooperating school systems. Level tests being constructed for math (Fall, 1976) and reading (Fall, 1977). When completed, should be possible to administer short test appropriate to "functioning level" of student and to secure more reliable measure than from longer tests formerly used. Scores from any of these level tests should be comparable to those from any other, and statistically combinable. Results reported by goal as well as by total score.